MULTIPERIOD PREDICTIONS FROM STOCHASTIC DIFFERENCE EQUATIONS BY BAYESIAN METHODS

Gregory C. Chow

Econometric Research Program
Research Memorandum No. 123
April 1971

The research described in this paper was supported by NSF Grant GS 2799.

PRINCETON UNIVERSITY
Econometric Research Program
207 Dickinson Hall
Princeton, New Jersey

MULTIPERIOD PREDICTIONS FROM STOCHASTIC DIFFERENCE EQUATIONS BY BAYESIAN METHODS

Gregory C. Chow

1. INTRODUCTION

Existing methods of parameter estimation, be they versions of least squares, maximum likelihood, or Bayes, that have been applied to systems of dynamic econometric equations to produce forecasts were not designed for the purpose of forecasting. In this paper, it is argued that these estimation methods may be inadequate if the resulting estimates are to be used to make ex ante predictions for more than one period ahead, and if the accuracy of the predictions is measured, as it usually is, by the mean squared errors. A formulation of the multiperiod prediction problem is presented. It will then become clear that the same set of parameter estimates cannot be optimal in making predictions for different time periods into the future, when optimality is defined by minimum mean squared errors in small samples.

Recently there has been much interest in comparing the ex post forecasting performance of different econometric models, e.g. Hickman [4], and of different estimation techniques applied to the same econometric model, e.g., Klein [8] and Fair [3]. It has become generally recognized that, in terms of ex post prediction tests, models or techniques that perform better for one-period predictions may do worse for multi-period predictions.

If one wished to have a model perform well in ex post predictions for many periods ahead (well in the sense of small squared prediction errors), one could fit the data with such a criterion in mind. A related, though different, question naturally arises as to whether different estimators should be used for making ex ante forecasts for different periods into the future. The former topic is one of fitting equations to a set of data. The latter topic is one of statistical decision, and is the subject of this paper.[1]

Section 2 treats the case of multiperiod predictions for a univariate time series which satisfies a first-order autoregression. A completely analytical solution to the prediction problem is given by the Bayesian method. The result of this section could be considered complementary to the recent work of Orcutt and Winokur [10] on first-order autoregression. Section 3 generalizes the result of section 2 to the case of a higher order system of linear stochastic difference equations with exogenous variables. It provides an alternative approach to multiperiod forecasting by the use of the reduced form of a system of linear econometric equations.

2. MULTI-PERIOD PREDICTIONS FROM A FIRST-ORDER AUTOREGRESSION

Consider the simplest case of a univariate first-order autoregression.

(2.1) $$y_t = a\, y_{t-1} + c + u_t \qquad (Eu_t = 0;\ Eu_t u_s = \delta_{ts}\, h^{-1}).$$

The intial value y_0 is treated as fixed, and u_t is assumed to be normally distributed. The parameters are the coefficients $\pi' = (a\ c)$ and the variance h^{-1} of the residuals. The problem is that, having observed $y' = (y_1, \ldots, y_n)$, one wishes to predict y_{n+k}. For the predictor $\hat{y}_{n+k} = \hat{y}_{n+k}(y)$ of y_{n+k}, assume the risk function to be

(2.2) $$R(\pi, h) = E(\hat{y}_{n+k} - y_{n+k})^2$$

$$= \int (\hat{y}_{n+k} - y_{n+k})^2\, p(y, y_{n+k} | \pi, h)\, dy\, dy_{n+k}$$

where, as elsewhere in this paper, the symbol p denotes probability density function which is assumed to exist. We will take the Bayesian approach of choosing the prediction function $\hat{y}_{n+k}(y)$ to minimize expected risk, given a prior density $p(\pi, h)$ of the parameters.

It will simplify the derivations to rewrite the risk (2.2) as an expectation taken over the future $u_{n+1}, \ldots, u_{n+k}$, rather than the future y_{n+k}. By the model (2.1), and on repeated substitutions of u's for y's, we have

$$(2.3) \quad y_{n+k} = a^k y_n + (a^{k-1}c + a^{k-2}c + \ldots + c) + (a^{k-1}u_{n+1} + a^{k-2}u_{n+2} + \ldots + u_{n+k}) .$$

Then, noting the fact that the future u's are uncorrelated with the observed y, or with the parameters, we can write the risk (2.2) as

$$(2.4) \quad E(\hat{y}_{n+k} - y_{n+k})^2$$

$$= E[\hat{y}_{n+k} - a^k y_n - (a^{k-1}c + \ldots + c)]^2 + E[a^{k-1}u_{n+1} + \ldots + u_{n+k}]^2 .$$

Since the prediction function $\hat{y}_{n+k}(y)$ affects only the first term on the right side of (2.4), the Bayesian approach amounts to minimizing, with respect to $\hat{y}_{n+k}$, the expression

$$(2.5) \quad \int E[\hat{y}_{n+k} - a^k y_n - (a^{k-1}c + \ldots + c)]^2 \, p(\pi, h) \, d\pi \, dh$$

$$= \iint [\hat{y}_{n+k} - a^k y_n - (a^{k-1}c + \ldots + c)]^2 \, p(y|\pi, h) \, dy \cdot p(\pi, h) \, d\pi \, dh .$$

This minimization will be achieved by minimizing

$$(2.6) \quad \int [\hat{y}_{n+k} - a^k y_n - (a^{k-1}c+\ldots+c)]^2 \, p(y|\pi,h) \cdot p(\pi,h) d\pi \, dh \; .$$

Setting the derivative of (2.6) with respect to $\hat{y}_{n+k}$ equal to zero, one obtains

$$(2.7) \quad \hat{y}_{n+k} = \frac{1}{\int p(y|\pi,h)\, p(\pi,h) d\pi \, dh} \int [a^k y_n + (a^{k-1}c+\ldots+c)] \cdot p(y|\pi,h)\, p(\pi,h)\, d\pi \, dh$$

$$= [\int a^k \, p(\pi|y) d\pi] y_n + \int (a^{k-1}c+\ldots+c) p(\pi|y) d\pi \quad .$$

Thus the optimal predictor $\hat{y}_{n+k}$ consists of two parts, the first being the expectation of a^k times the initial value y_n, and the second being the expectation of $(a^{k-1}c + a^{k-2}c+..+c)$, where the expectations are evaluated by the posterior density $p(a,c|y) = p(\pi|y)$. If one had formulated a similar Bayesian problem of estimating a and c , and not of predicting y_{n+k} , his point estimates $\hat{a}$ and $\hat{c}$ would be the expectations of the posterior density $p(a,c|y)$. A predictor constructed from these estimates would be $\hat{a}^k y_n + (\hat{a}^{k-1} \hat{c} +\ldots+ \hat{c})$, and would not in general be optimal by the criterion of minimum mean squared error. The two predictors would coincide only for the prediction of y_{n+1} .

It is well-known that if the density $p(y|\pi,h)$ is normal, and the prior density $p(\pi,h)$ is either diffused, being proportional to h^{-1} as suggested by Jeffreys [5], or is normal-gamma as suggested by Raiffa and Schlaifer [11], the posterior density is normal-gamma. Under the assumption of the diffused prior, the posterior density is

$$(2.8) \qquad p(\pi,h|y) \propto h^{\frac{1}{2}n-1} \exp\{-\tfrac{1}{2}h(y-Z\pi)'(y-Z\pi)\}$$

where

$$(2.9) \qquad y' = [y_1 \cdots y_n] \; ; \quad Z' = \begin{bmatrix} y_0 \cdots y_{n-1} \\ 1 \;\cdots 1 \end{bmatrix} ; \quad \pi' = [a \;\; c] \; .$$

On integrating over h, the marginal posterior density of π is well-known to be the bivariate t density

$$(2.10) \qquad p(\pi|y) \propto [(y-Z\pi)'(y-Z\pi)]^{-\frac{1}{2}n}$$

$$\propto [(y-Z\hat{\pi})'(y-Z\hat{\pi}) + (\hat{\pi}-\pi)'Z'Z(\hat{\pi}-\pi)]^{-\frac{1}{2}n}$$

$$\propto [(n-2) + (\pi-\hat{\pi})'(\frac{1}{s^2}Z'Z)(\pi-\hat{\pi})]^{-\frac{1}{2}n}$$

where

$$(2.11) \qquad \hat{\pi} = \begin{bmatrix} \hat{a} \\ \hat{c} \end{bmatrix} = (Z'Z)^{-1}Z'y \; ; \quad (n-2)s^2 = (y-Z\hat{\pi})'(y-Z\hat{\pi}) \; .$$

The posterior density (2.10) can be applied to equation (2.7) to obtain optimal predictions $\hat{y}_{n+k}$. For k=1,

$$(2.12) \qquad \hat{y}_{n+1} = [\int a\, p(\pi|y)d\pi]\, y_n + \int c\, p(\pi|y)d\pi$$

$$= (Ea)y_n + (Ec) = \hat{a}\, y_n + \hat{c} ,$$

where the expectation E is taken over the posterior density (2.10). For k=2 ,

$$(2.13) \qquad \hat{y}_{n+2} = (Ea^2)y_n + (E\, a\, c) + (Ec) .$$

To evaluate the second moments, we note

$$(2.14) \qquad E\begin{bmatrix} a^2 & ac \\ ac & c^2 \end{bmatrix} = E\begin{bmatrix} a - \hat{a} \\ c - \hat{c} \end{bmatrix} [a-\hat{a} \quad c-\hat{c}] + \begin{bmatrix} \hat{a} \\ \hat{c} \end{bmatrix} [\hat{a} \quad \hat{c}]$$

$$= \frac{n-2}{n-4}\, s^2 (Z'Z)^{-1} + \begin{bmatrix} \hat{a}^2 & \hat{a}\,\hat{c} \\ \hat{c}\,\hat{a} & \hat{c}^2 \end{bmatrix}$$

where the second moments around the means for the multivariate t density can be found in Raiffa and Schlaifer [11, p. 257]. Thus the predictor (2.13) differs from the usual one by

$$(2.15)\quad [Ea^2 - \hat{a}^2]y_n + [Eac - \hat{a}\hat{c}] = \frac{(n-2)s^2}{n-4}[g_{11}y_n + g_{12}]$$

where g_{ij} is the i-j element of $(Z'Z)^{-1}$.

In general, the predictor $\hat{y}_{n+k}$ involves $E\,a^k$ and terms of the form $Ea^i c$. To evaluate $E\,a^k$, one uses the marginal density of a, which is univariate t given by [11,p.258]

$$(2.16)\quad p(a|y) \propto [\,(n-2) + s^2 g_{11}(a-\hat{a})^2]^{-\frac{1}{2}(n-1)}$$

Let $\mu_j = E(a-\hat{a})^j$. Then $\mu_j = 0$ for j odd; and, for j even, μ_j is given by [8, p. 60]

$$(2.17)\quad \mu_j = \frac{\Gamma(\frac{1}{2}j + \frac{1}{2})\,\Gamma(\frac{1}{2}n - \frac{1}{2}j - 1)}{\Gamma(\frac{1}{2})\,\Gamma(\frac{1}{2}n-1)} \qquad (n > j + 2) .$$

$E\,a^k$ can then be calculated by the formula [8, p. 56]

$$(2.18)\quad E\,a^k = \sum_{j=0}^{k} \binom{k}{j} \mu_{k-j}\,\hat{a}^j .$$

To evaluate $E\,a^i c$, one can use the conditional mean of c [11,p.258],

$$(2.19)\quad \int c\,p(c|a,y)\,dc = \hat{c} - g_{21}g_{11}^{-1}\hat{a} + g_{21}g_{11}^{-1}a ,$$

and then use

$$(2.20) \qquad E\, a^i c = \int a^i [\int c\, p(c|a,y)\, dc]\, p(a|y)\, da$$

$$= (\hat{c} - g_{21} g_{11}^{-1} \hat{a}) \int a^i p(a|y)\, da + g_{21} g_{11}^{-1} \int a^{i+1} p(a|y) da$$

where the means of a^i and a^{i+1} can be calculated as in the last paragraph.

Note that if predictions for $y_{n+1}, y_{n+2}, \ldots, y_{n+K}$ are required simultaneously, each can be constructed by the method of equation (2.7). This procedure can be justified by reformulating the risk function (2.2) to be $\sum_{k=1}^{K} E(\hat{y}_{n+k} - y_{n+k})^2$, and equation (2.7) for each value of k will follow from the same derivations as given above.

The result of this section complements those of Orcutt and Winokur [10] on first-order autoregression, where results from Monte Carlo experiments on the properties of some estimators and predictors were reported. It might be of interest to compare the multi-period predictors of this section with those examined by Orcutt and Winokur through further Monte Carlo experiments.

3. GENERALIZATION TO SYSTEMS OF STOCHASTIC DIFFERENCE EQUATIONS

Consider a (pxl) vector y_t obeying

$$(3.1) \quad y_t = A_1 y_{t-1} + \cdots + A_q y_{t-q} + C_o x_t + \cdots + C_r x_{t-r} + u_t \quad (E u_t u_s' = \delta_{t,s} H^{-1})$$

Again, $y_o, y_{-1}, \ldots, y_{-q+1}$ are treated as fixed, and u_t is assumed to be normal. The parameters are $\Pi' = (A_1 \cdots A_q; C_o \cdots C_r)$ and H. Having observed $Y' = (y_1 \cdots y_n)$ with $n > q + r + 1$, one wishes to predict y_{n+k} conditionally on all future values of the exogenous variables up to x_{n+k}. For the predictor $\hat{y}_{n+k}(Y)$ of y_{n+k}, assume the risk to be, with D symmetric and positive definite,

$$(3.2) \qquad R(\Pi, H) = E(\hat{y}_{n+k} - y_{n+k})' D (\hat{y}_{n+k} - y_{n+k}) .$$

To achieve a generalization of the result of section 2, we rewrite the model (3.1) as a first order system

$$(3.3) \quad \begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-q+1} \end{bmatrix} = \begin{bmatrix} A_1 & A_2 \cdots A_q \\ I & 0 \cdots 0 \\ & \cdots \\ 0 \cdots & I \;\; 0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-q} \end{bmatrix} + \begin{bmatrix} C_o \cdots C_r \\ 0 \cdots 0 \\ \cdots \\ 0 \quad\; 0 \end{bmatrix} \begin{bmatrix} x_t \\ \vdots \\ x_{t-r} \end{bmatrix} + \begin{bmatrix} u_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

or, in new symbols,

(3.4) $$y_t^* = A\, y_{t-1}^* + C\, x_t^* + u_t^* \ .$$

If $\hat{y}_{n+k}^*$ denotes the predictor of y_{n+k}^*, we are interested only in its upper-most subvector $\hat{y}_{n+k}$, but let us first use model (3.4) to express y_{n+k}^* in terms of the future u^*'s as in (2.3),

(3.5) $$y_{n+k}^* = A^k y_n^* + (A^{k-1} C x_{n+1}^* + \dots + C x_{n+k}^*) + (A^{k-1} u_{n+1}^* + \dots + u_{n+k}^*) \ .$$

The relevant subvector y_{n+k} is obtained as $M\, y_{n+k}^*$, where the matrix $M = (I_p \ 0 \cdots 0)$ selects the first p rows of y_{n+k}^*. Using (3.5), the definition of M, and the independence of the future u^*'s, we write the risk function (3.2) as

(3.6) $$E\|\hat{y}_{n+k} - y_{n+k}\|_D^2 = E\|\hat{y}_{n+k} - M\, y_{n+k}^*\|_D^2$$

$$= E\|\hat{y}_{n+k} - M A^k y_n^* - M(A^{k-1} C x_{n+1}^* + A^{k-2} C x_{n+2}^* + \dots + C x_{n+k}^*)\|_D^2$$

$$+ \ E\|M(A^{k-1} u_{n+1}^* + \dots + u_{n+k}^*)\|_D^2$$

where, the term involving future u^*'s is not affected by the predictor $\hat{y}_{n+k}$.

The development from (2.5) to (2.7) can be followed exactly to yield the result

$$(3.7)\quad \hat{y}_{n+k} = [\int MA^{k}\, p(\Pi|Y)d\Pi]y^{*}_{n} + [\int MA^{k-1}C p(\Pi|Y)d\Pi]x^{*}_{n+1} + \dots + [\int MC p(\Pi|Y)d\Pi]x^{*}_{n+k} \quad .$$

The interpretation of (3.7) is similar to that of (2.7). Rather than first obtaining the estimate $\hat{\Pi}' = (\hat{A}_1,\dots,\hat{A}_q;\ \hat{C}_o\dots\hat{C}_r)$ of Π' as the mean of the posterior density and then form the predictor of y_{n+k} as $M\hat{A}^{k}y^{*}_{n} + M\hat{A}^{k-1}\hat{C}\, x^{*}_{n+1} + \dots + M\hat{C}\, x^{*}_{n+k}$, one should compute the expectations of MA^{k}, $MA^{k-1}C,\dots,MC$ by the posterior density, and then apply them to y^{*}_{n}, $x^{*}_{n+1},\dots,x^{*}_{n+k}$ respectively. Again, the two methods would give the same result only for the prediction of y_{n+1}.

It is well-known that if the density $p(y|\Pi,H)$ is normal, and the prior density $p(\Pi,H)$ is either diffused, being proportional to $|H|^{-\frac{1}{2}(p+1)}$, or is normal-Wishart, the posterior density is normal-Wishart. Under the assumption of the above diffused prior, the posterior density is

$$(3.8)\quad p(\Pi,H|Y) \propto |H|^{\frac{1}{2}(n-p-1)} \exp\{-\tfrac{1}{2}\,\mathrm{tr}\, H\,(Y-Z\Pi)'(Y-Z\Pi)\}$$

where Y is $n\times p$, Z is $n\times s$ $(s=q+r+1)$, and Π is $s\times p$, as defined by

$$(3.9)\quad Y' = [y_1 \dots y_n]\ ;\quad Z' = \begin{bmatrix} y_0 & & y_{n-1} \\ \vdots & \vdots & \vdots \\ y_{1-q} & \cdot & y_{n-q} \\ x_1 & & x_n \\ \vdots & & \vdots \\ x_r & & x_{n-r} \end{bmatrix}\ ;\ \Pi' = [A_1 \dots C_r]\ .$$

On integrating over H, using the integral of the Wishart distribution [1, p. 154], one obtains the marginal posterior density of Π to be the matricvariate t density[2]

$$(3.10)\quad p(\Pi|Y) \propto |(Y - Z\Pi)'(Y - Z\Pi)|^{-\frac{1}{2}n}$$

$$\propto |(Y - Z\hat{\Pi})'(Y - Z\hat{\Pi}) + (\Pi - \hat{\Pi})'Z'Z(\Pi - \hat{\Pi})|^{-\frac{1}{2}n}$$

where $\hat{\Pi}$ is $(Z'Z)^{-1}Z'Y$. The posterior density (3.10) can be applied to equation (3.7) for obtaining the optimal prediction $\hat{y}_{n+k}$.

While the problem of numerical integration of (3.7) remains to be further investigated, the following analytical results may be used at least for evaluating predictions when k is small. To evaluate $\hat{y}_{n+k}$ by (3.7), it is required to calculate expectations of products of at most k elements of Π. Let the p columns of Π be arranged into a column vector $\pi = (\pi_1 \dots \pi_{ps})'$ of ps elements. To calculate the expectation of $\pi_i\pi_j\dots\pi_m$, one can use the posterior density $p(\Pi,H|Y)$ and integrate first over Π:

$$(3.11)\quad \int(\pi_i\pi_j\ldots\pi_m)\,p(\Pi,H|Y)\;d\Pi\cdot dH$$

$$\propto \int |H|^{\frac{1}{2}(n-p-1-s)}\Big[\int(\pi_i\ldots\pi_m)|H|^{\frac{1}{2}s}\exp\{-\tfrac{1}{2}\,\mathrm{tr}\,H(\Pi-\hat{\Pi})'Z'Z(\Pi-\hat{\Pi})\}\,d\Pi\Big]$$

$$\cdot\ \exp\{-\tfrac{1}{2}\,\mathrm{tr}\,H(Y-Z\hat{\Pi})'(Y-Z\hat{\Pi})\}\,dH$$

where the density of Π, given H, is seen to be multivariate normal with mean $\hat{\Pi}$ and the covariance matrix of the vector π is $H^{-1}\otimes(Z'Z)^{-1} \equiv V = (v_{ij})$. The expectation of the product $\pi_i\pi_j\ldots\pi_m$ of k π's is known, as it can easily be derived from the characteristic function of the multivariate normal density, to be $\hat{\pi}_i\hat{\pi}_j\ldots\hat{\pi}_m$, plus products of one v_{ij} and $(k-2)$ $\hat{\pi}$'s, plus products of two v_{ij} and $(k-4)$ $\hat{\pi}$'s, etc., the k subscripts in each product being a permutation of $(i,j,\ldots,m)$ but with v_{ij} and v_{ji} counted only once.[3] Thus for $k = 2,3$ at most one v_{ij} is involved in each term; for $k=4,5$ the product of at most two v_{ij} is involved in each term. The former involves one element of H^{-1}, and the latter involves the product of at most two elements of H^{-1}.

Having found the expectation of the product $\pi_i\ldots\pi_m$ given H, one has to perform the remaining integration of (3.11) over H. The remaining density of H in (3.11) is the Wishart density with parameter set $[(Y-Z\hat{\Pi})'(Y-Z\hat{\Pi}),\ (n-s)]$. It is thus required to find the expectation of an element (for $k=2,3$), or of the product of two elements (for k=4,5), of H^{-1}, as the $\hat{\pi}$'s and the elements of the matrix $(Z'Z)^{-1}$ are given for the purpose of this integration. Let h^{ij} be the i-j element of the pxp matrix H^{-1}, s_{ij} be the i-j element of $(Y-Z\hat{\Pi})'(Y-Z\hat{\Pi})$, and let $\nu = (n-s)-p+1$. Kaufman [6, p. 14] has obtained

$$(3.12) \qquad E\, h^{ij} = \frac{1}{\nu-2} s_{ij} \quad ;$$

$$E(h^{ij}h^{k\ell}) = \frac{1}{(\nu-1)(\nu-2)(\nu-4)} [(\nu-3)s_{ij}s_{k\ell} + s_{ik}s_{j\ell} + s_{i\ell}s_{kj}] \; .$$

These results complete the ahalytical solutions to the prediction of y_{n+k} for k up to 5.

Insofar as the reduced form equations in linear econometric models are systems of stochastic difference equations, the result of this section applies to multiperiod predictions by econometric models. While this paper has brought out the dynamic aspect of forecasting, it has ignored the non-linear restrictions on reduced-form coefficients induced by over-identification of the structural equations, an aspect emphasized in the existing literature on the estimation of sumultaneous equation systems. Which aspect of the problem will turn out to be more important for forecasting purposes remains an open question.[4]

REFERENCES

[1] Anderson, T.W.: Introduction to Multivariate Statistical Analysis, (New York: John Wiley and Sons, Inc., 1958).

[2] Drèze, J.H. and Morales, J.A.: "Bayesian Full Information Analysis of the Simultaneous Equations Model," Center for Operations Research and Econometrics Discussion Paper No. 7031, Catholic University of Louvain (August 1970).

[3] Fair, R.C.: "A Comparison of Alternative Estimators of Macroeconomic Models," Econometric Research Program Research Memorandum No. 121 (March 1971).

[4] Hickman, B.G.: Ed., Econometric Models of Cyclical Behavior, (National Bureau of Economic Research, New York, to be published).

[5] Jeffreys, H.: Theory of Probability, (Oxford: Clarendon Press, 1948).

[6] Kaufman, G.M.: "Some Bayesian Moment Formulae," Center for Operations Research and Econometrics Discussion Paper No. 6710, Catholic University of Louvain (August, 1967).

[7] Kaufman, G.M.: "Conditional Prediction and Unbiasedness in Structural Equations," Econometrica, Vol. 37, No. 1, (January, 1969), pp. 44-49.

[8] Kendall, M.G., and Stuart, A.: The Advanced Theory of Statistics, Vol. I (London: Charles Griffin and Co., 1958).

[9] Klein, L.R.: "Estimation of Interdependent Systems in Macro-econometrics," Econometrica, Vol. 37, No.2 (April, 1969), pp. 171-192.

[10] Orcutt, G.H. and Winokur, H.S., Jr.: "First Order Autoregression: Inference, Estimation, and Prediction," Econometrica, Vol. 37, No. 1 (January, 1969), pp. 1-14.

[11] Raiffa, H. and Schlaifer, R.: Applied Statistical Decision Theory, (Boston: Harvard University Graduate School of Business Administration, 1961).

[12] Sewell, W.P.: "Least Squares, Conditional Predictions, and Estimator Properties," Econometrica, Vol. 37, No. 1, (January, 1969), pp. 39-43.

[13] Zellner, A. and Chetty, V.K.: "Prediction and Decision Problems in Regression Models from the Bayesian Point of View," Journal of American Statistical Association, Vol. 60, No. 310, (June, 1965), pp. 608-616.

FOOTNOTES

1. I should like to acknowledge that my interest in the subject of this paper has partly stemmed from conversations with Ray C. Fair in connection with his work on the former subject [3].

2. Recent applications of the matricvariate t density to econometric problems include papers by Zellner and Chetty [13], Sewell [12], Kaufman [7], and Dreze and Morales [2], among others. None of these studies deals with the multi-period prediction problem of the present paper, however.

3. Thus for $k=5$, we have

$$E\pi_i\pi_j\pi_k\pi_\ell\pi_m = \hat{\pi}_i\hat{\pi}_j\hat{\pi}_k\hat{\pi}_\ell\hat{\pi}_m + \binom{5}{2} \text{ terms of the form } v_{ij}\hat{\pi}_k\hat{\pi}_\ell\hat{\pi}_m$$

$$+ \tfrac{1}{2}\binom{5}{2}\binom{3}{2} \text{ terms of the form } v_{ij}v_{k\ell}\hat{\pi}_m \; .$$

4. Of course, both aspects could be incorporated, at least in principle. One way to do this would be to utilize the posterior density $p(B,\Gamma|Y)$ of Dreze and Morale [2] for the coefficients B and Γ in the structure $B'y_t = \Gamma'z_t$ + residuals, where a priori restrictions on the elements of B and Γ have been imposed. Our solution would require evaluating elements of the integral

$$\int \Pi^k \, p(\Pi|Y)d\Pi = \int(\Gamma B^{-1})^k \, p(B,\Gamma|Y)d(B,\Gamma) \quad .$$

I have not examined the feasibility of this numerical problem.